Chapter 3

Procedure for developing a thermodynamic metabolic model

Bastian Niebel, Simeon Leupold, Matthias Heinemann

(Manuscript in preparation)

Recently, we developed a new flux balance analysis-based approach, which enabled us to predict extracellular and intracellular fluxes of *Saccharomyces cerevisiae* with unprecedented accuracy. These predictions were performed with a thermodynamic metabolic network model including a comprehensive description of the biochemical thermodynamics: the second law of thermodynamic, a cellular Gibbs energy balance, and an upper limit on the cellular entropy production rate. Here, we present a workflow that outlines the steps to build such a thermodynamic metabolic network model from an existing metabolic reconstruction, and using thermodynamic data for the metabolic processes and a training data set. This workflow consists of the four steps, (I) adding biochemical information to the stoichiometric network, (II) determining Gibbs energies for the metabolic processes, (III) reducing the stoichiometric network, and (IV) training the model on experimental data. Here, in addition to the model for *S. cerevisiae*, which is described in Chapter 2, we apply steps (I-II) of this workflow on the most recent genome scale reconstruction for *Escherichia coli*.

BN and SL contributed equally to this work. BN developed the procedure. SL carried out the computational simulations and analyzed the data. BN and SL made the figures. BN wrote the manuscript. SL contributed to the writing. MH supervised the research and contributed to the writing

Introduction

In the past decades, metabolic networks of different organisms, ranging from bacteria, fungi, mammalian cells, to plants, have been reconstructed (126-129). These networks consist of the steady-state mass balances for the metabolites and describe the rates through metabolic processes, i.e. chemical conversions and metabolite transport. Metabolic network models have become a valuable tool in research and for applications (4, 6).

An often used tool to analyze these models is flux balance analysis (FBA) (14). FBA uses mathematical optimization to address the problem that these network models represent an underdetermined system of equations. Typically, an objective function, such as maximizing the cellular growth rate subject to the constraints defined by the mass balances of the metabolic network, is optimized. While FBA has been found to generate reasonable predictions in some conditions (4, 6, 14), so far no condition-independent set of constraints and objective functions could be identified that were able to correctly predict fluxes across different conditions (130).

Recently, we identified a thermodynamic constraint (Chapter 2), which, when implemented into a thermodynamic metabolic network model for *Saccharomyces cerevisiae* and used with the objective of biomass maximization, gave correct FBA-predictions across a wide range of different conditions. This identified constraint enforces that metabolic operation does not exceed a maximum rate of cellular entropy production. The maximal rate has to be determined for the given organism from experimental data, i.e. physiological rates and metabolite levels. Then it can be used with a stoichiometric metabolic network model including a correct description of the biochemical thermodynamics to predict metabolic fluxes accurately under wide range of conditions.

A metabolic network model with a comprehensive description of the biochemical thermodynamics, i.e. a thermodynamic metabolic network model, contains (i) mass-, charge- and pH-dependent-proton balances, (ii) the second law of thermodynamics, and (iii) a cellular Gibbs energy balance (Chapter 2, methods 2), which ties the rates of Gibbs energy exchange with the environments to the cellular rate of entropy production, σ^{cell}, where later rate is the sum of the metabolic processes' entropy production rates.

Here, we present a workflow consisting of four steps, with which a genome-scale model can be translated into a thermodynamic metabolic network model. This workflow was developed based on the model building process of Chapter 2, which resulted in a model for *S. cerevisiae*. We illustrate the first three steps on this workflow on the most recent genome scale reconstruction for *Escherichia coli*.

Workflow

In order to use FBA with an upper limit in the cellular entropy production rate, $\sigma^{cell}{}_{up}$-constrained FBA, a metabolic network model needs to be augmented for a proper thermodynamic description and constraints. Such a thermodynamic metabolic network model $M(v,\ln c) \leq 0$ (Chapter 2, methods 5) consists of a system of linear and non-linear equations with the variables denoting the rates v through the metabolic processes and

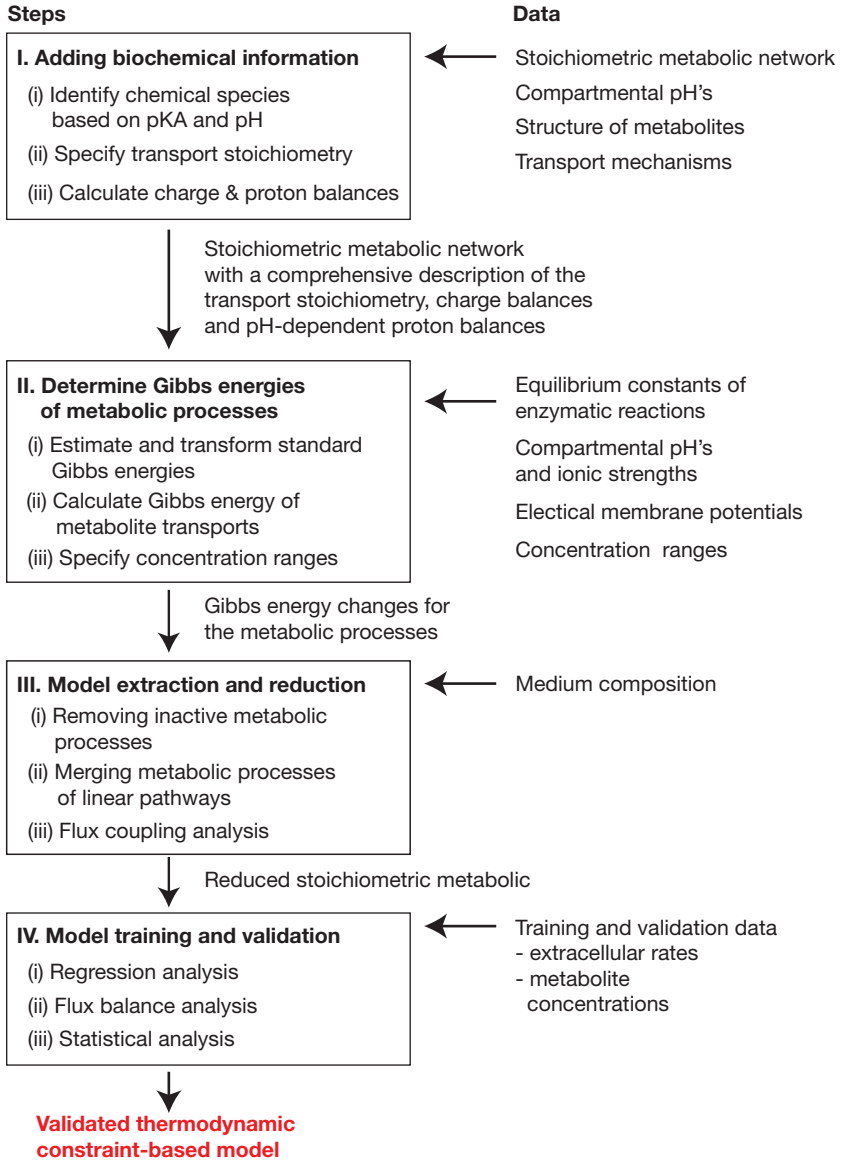

Figure 1. Workflow diagram for the development of a thermodynamic metabolic network model in four steps.

the metabolite concentrations ln c (on a logarithmic scale). The solution space of the thermodynamic metabolic network model defines the possible metabolic operations. In order to define $M(v,\ln c) \leq 0$, the following data are necessary: the stoichiometric coefficients for the reactions/transporters that define the metabolic network, Gibbs energies changes of the metabolic processes, and rough condition-independent bounds on the metabolite concentrations, which can be obtained from a collection of quantitative metabolome measurements.

Here, we present a workflow (Fig. 1) to build such a thermodynamic metabolic network model starting from a genome-scale metabolic reconstruction (131-134). This workflow consists of four steps: (I) adding biochemical information to the stoichiometric network, (II) determining the Gibbs energy changes of the metabolic processes, (III) reducing the stoichiometric network, and (IV) training the thermodynamic metabolic network model in order to identify the unknown model parameters. Besides the model for *S. cerevisiae* (refer for the specific details on the implementation to Chapter 2, Extended Methods, which has been tested and validated), we illustrate the steps (I-III) with the concrete example of converting a recent genome-scale model of *E. coli* into a thermodynamic metabolic model.

Step I. Adding biochemical information to the stoichiometric network

Typically, genome-scale metabolic reconstructions contain a-priori defined proton balances and - in order to obtain biological meaningful results - constraints on the directionality of certain metabolic processes (57). In a first step, we remove the a-priori proton balances from the metabolic network, because in our thermodynamic metabolic network models we use pH-dependent proton balances (Chapter 2, methods 1). Further, in order to remove any heuristic assumptions, in our model, we constrain the directionality of all metabolic processes only by the 2nd law of thermodynamics, and therefore we also remove in a preprocessing step, the directionality assignments from the metabolic network. Also, we merge the periplasmic and the extracellular space into one compartment, since the selectivity of the outer membrane is rather low, and chemical conditions (namely pH and ionic strength) are similar (135), and therefore the transport over the outer membrane can be neglected.

After these preprocessing steps, we then (i) determine for each metabolite the chemical species that are present at the compartments' pH value, (ii) add metabolite transport processes to allow translocation of individual chemical species, and (iii) add charge and pH-dependent proton balances to enforce pH- and charge homeostasis in each compartment, where these balances are influenced by the chemical conversions and transport of chemical species. Details of these steps are illustrated in the following paragraphs.

Note, we differentiate between the reactants, i.e. the equilibrated mixture of chemical species at a specific pH, and chemical species, because this differentiation allows us to model the transport of chemical species between cellular compartments at different pH values (Fig. 2A) (35). The differentiation between chemical species and reactants in the charge balances and pH-dependent proton balances and metabolite transport allows to correctly model energy transfers between different metabolic processes.

(i) The chemical species of each metabolite (reactant) are determined from the acid dissociation constants (pKa) of the metabolite. The pKa values are estimated from the chemical structure of the metabolites, using the Marvin Suit from ChemAxon. This estimation is based on

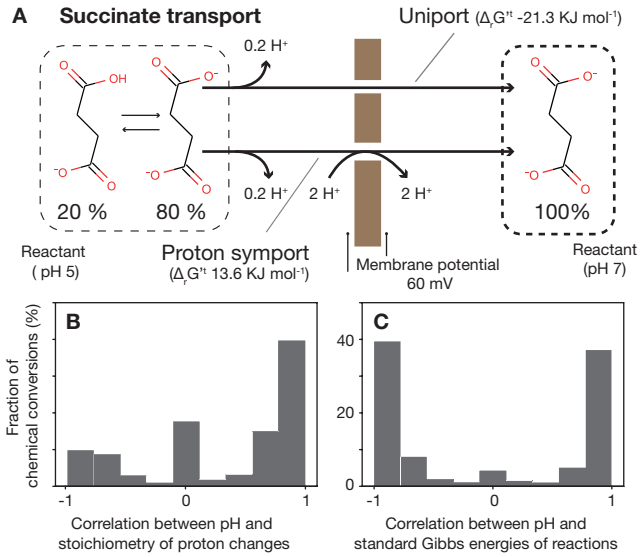

Figure 2. Adding biochemical information to the metabolic network model. **(A)** Example of a transport process and its modelling (transport of unprotonated succinate), where we included two different variants (uniport and proton symport) of this transport process to take into account differences in the charge balance, proton balances, and the Gibbs energy of this metabolite transport (For the mathematical model description of the transport processes refer to Chapter 2, methods 1 & 3) **(B)** Histogram of Spearman rank correlation coefficients between pH values and stoichiometric coefficients describing the changes in protons of 1473 reactions shows the pH-dependence of the proton balances. **(C)** Histogram of Spearman rank correlation coefficients between pH values and standard transformed Gibbs energies of 1473 reactions shows the sensitivity of these Gibbs energies with respect to changes in pH.

the partial charge distribution of each molecule and a set of training molecules. The chemical structure of the metabolites is downloaded from the KEGG database (136) and and encoded as IUPAC International Chemical Identifier (InChI) (137). Based on the pKa values and the compartmental pH values, we then calculate the abundance fraction of each chemical species of a metabolite (reactant) (138). In order to keep the model compact, we only consider chemical species, which have a fractional abundance of at least 10%. Charge and amount of protons for species of metabolites, whose structures cannot be properly encoded as InChI, such as molecular clusters, e.g. Fe-S clusters, or complex organometallics, were manually defined based on structural knowledge and databases such as EcoCyc (139) or ecmdb (140). For metabolites with unknown or ambiguous molecule structures, such as proteins or tRNA, where pKA values cannot be determined, we only considered one species with neutral charge.

In this step, it is advisable to check the elemental balances (for all elements except hydrogen) of all chemical transformations in the metabolic network. This consistency check detects potential discrepancies between the metabolites' elemental compositions specified by (a) the InChI and (b) the original genome-scale metabolic network. If there is a discrepancy, the chemical transformations should be corrected to account for the elemental composition specified by the InChI.

(ii) As for most metabolite transporters the exactly transported species is not known, we introduce for every metabolite transport multiple variants in order to allow for transport of different chemical species:

For metabolites that are transported by diffusion, proton symporter, proton antiporter, or unknown transport mechanism, we allow the transport of the species (with fractional abundance greater than 10%) or species combinations. Additionally, in case of transport processes, for which at the given pH value no charge neutral transport variant exists, we introduce an additional transport reaction, in which protons—balancing the charge—are co-translocated together with the respective species, i.e. adding a proton symporter or antiporter. This additional transport variant ensures that for every metabolite a transport variant exists that does not translocate net charge. For metabolites that are transported by active transport mechanisms, we include transport processes for all considered species (with fractional abundance greater than 10%), where we accurately model the transport mechanism, e.g. for the phosphotransferase system, first the unphosphorylated molecule is transported over the membrane and subsequently phosphorylated in the cytoplasmic space.

For redox reactions, where half reactions are taking place on both sides of a membrane, e.g. complexes of the respiratory chain, we also model the transfer of electrons across the membrane, since this electron transfer influences the charge and Gibbs energy changes of the transport process (for details refer to Chapter 2, methods 1, 3). To model these redox processes, one needs to determine, on which side of the membrane the redox cofactors are located and needs to add a chemical species for the electrons. The location of the redox cofactors are determined from the mechanisms of this membrane-bound redox reactions. The necessary information has to be taken from biochemical textbooks and the scientific literature. For an example please refer to the implementation of the respiratory chain of the *S. cerevisiae* model (Chapter 2, fig. S1, table S5-S7).

(iii) To specify the charge and proton balances, we calculate the stoichiometric coefficients of the charge and proton changes associated with the metabolic processes. The stoichiometric coefficients of the charge changes are based on the charges of the translocated chemical species and can be calculated according to Chapter 2, methods 1. Note, in case the metabolic network does not account for ions, e.g. potassium, calcium or magnesium, we introduce a generic ion species, for which we also include different transport variants, i.e. a uniport, a symport, and an ATP-driven active transport. Because the number of hydrogen atoms of reactants depends on the compartmental pH values, the stoichiometric coefficients of the proton changes are pH-dependent (Fig 2, A and B). To take into account this pH-dependency, the stoichiometric coefficients of the proton changes are calculated according to Chapter 2, methods 1, where we need the fractional abundances of the chemical species in the reactants (see (i)).

For building the thermodynamic metabolic network model of *E. coli*, the stoichiometry of the metabolic network was gathered from the most recent genome-scale metabolic reconstruction network (126), comprising 1136 unique metabolites located in the cytosol, periplasmatic and the extracellular space, 1471 chemical conversions, 782 metabolite transporter and 330 exchange processes. By merging the periplasmatic and the extracellular space, 325 metabolite transport processes between both compartments were removed from the model and 193 periplasmatic chemical conversions were relocated into the extracellular space.

Subsequently, the remaining set of 457 metabolite transport processes were enlarged to 604 transport variants by accounting for the different species (with fractional abundance greater than 10%). The resulting model contains in total 1132 metabolites and 2075 metabolic processes (1471 chemical conversions, 604 metabolite transport) and 330 exchange processes.

Step II. Determining Gibbs energies changes of metabolic processes

Next, we need to gather the necessary data to calculate the Gibbs energies of the metabolic processes (Chapter 2, methods 3). These Gibbs energies are used in the thermodynamic metabolic network model to constrain the directionality of the metabolic processes through the second law of thermodynamics and to calculate the entropy production rates. To calculate the Gibbs energies of metabolic processes, we need (i) standard Gibbs energies of reactions, which describe the differences in the reactants' Gibbs energies of formations, (ii) Gibbs energies of metabolite transports associated with pH gradients, with translocation of protons by proton pumps, and with transport of charged metabolites over membrane potentials (Fig. 2A), and (iii) default ranges for the metabolite concentrations spanning typical physiological concentration levels. In the following paragraphs, we further illustrate these three different kinds of input data.

(i) We estimate the standard Gibbs energies of formations and reactions using the component contribution method (CCM) (27). CCM infers the standard Gibbs energies based on measured equilibrium constants (31), and approximations for the Gibbs energies by the group contribution (29). The CCM estimates the means and errors for the standard Gibbs energies of reactions. Both estimates are later used to determine a thermodynamically consistent set of Gibbs reaction energies (cf. section 'train the thermodynamic metabolic model' (Step IV)). The Gibbs energy of formation of the biomass is estimated based on the elemental biomass composition using an empirical formula (69). Because metabolic processes take place in cellular compartments with different pH values and ionic strengths, we take into account the effect of pH and ionic strengths on the standard Gibbs energies of reactions. We do this, because the results of a sensitivity analysis for all 1473 chemical conversion show that this is indeed necessary (Fig. 2C). Therefore, the standard Gibbs energies are transformed to the compartmental pH values and ionic strengths (21). The Gibbs energy of formation of the biomass is transformed to the cytosolic pH.

(ii) The Gibbs energies of metabolite transports are calculated using the formula described in Chapter 2, methods 3. Here, we need the fractional abundances of the chemical species in the reactants (see Step I), and electrical potentials across cellular membranes, which can be found in the scientific literature.

(iii) Typically, intracellular metabolite concentrations vary between 1 μM and 10 mM (36-39), which we use as default range. For certain metabolites, such as central carbon metabolism intermediates, amino acids, redox cofactors, oxygen or carbon dioxide, we use ranges, which are based on minimal and maximum values reported in scientific literature or their maximum solubility (141,142). These data were reprocessed using a condition independent dry weight specific cellular volume of 0.0023 L g_{DW}^{-1} (141).

For building the thermodynamic metabolic network model of *E. coli*, we determined the Gibbs energies based on a pH of 7.6 (143) in the cytosol and 7.0 in the extracellular space

(corresponding to M9 medium). The ionic strength was assumed to be 0.15 M (144) in the cytosol and 0.2 M in the extracellular space (corresponding to M9 medium). Based on the component contribution method, we were able to determine the Gibbs energy of reactions for 73% of the chemical transformations. The Gibbs energy of reactions, for which the CCM was not able to make an estimate, was allowed to vary by $\pm 10^5$ KJ/mol.

Step III. Reducing the stoichiometric network

Third, we reduce the metabolic network (example given in Fig 3A) without compromising its predictive power. Because in the model every metabolic processes is related to a binary variable (by the second law of thermodynamics) and a non-linear equation (by the entropy production rate), a reduction in the number of metabolic processes speeds up the computational analyses of the model. Such reduction step is advisable to facilitate the later computational analysis of the thermodynamic metabolic model, but is not mandatory to proceed with step III. A reduction can be performed by specifying the scope of the model via defining possible growth substrates and products by constraining the extracellular rates of these compounds accordingly. After such a specification, a number of different model reductions are possible: (i) identifying inactive metabolic processes, i.e. metabolic processes that can never carry metabolic flux at the specified conditions, (ii) merging of reactions within a linear pathway, and (iii) identifying linear dependent processes, i.e. reactions that have fully coupled fluxes (Note, step (iii) is not a reduction in the model size, but a reduction in the model complexity to facilitate later computational analyses).

(i) First, we identify inactive metabolic processes (Fig. 3B), using flux variability analysis (16), where we minimize and maximize the flux through each metabolic process using the mass balances as constraints. If the minimum and the maximum flux of a metabolic process is zero, then the metabolic process is inactive for the selected scope and can be removed from the network.

(ii) Second, the metabolic processes of linear pathways can be merged into one lumped metabolic process (Fig. 3C). This is possible, since - under steady-state conditions—the metabolic fluxes through reactions of a linear pathway are equal. Here, we identify linear pathways from the structure of the stoichiometric matrix including the charge and proton balances, by identifying rows in the stoichiometric matrix that only contain two non-zero stoichiometric coefficients, which correspond to metabolites with only one production and consumption process. We can then eliminate this metabolite by merging its producing and consuming metabolic process into one lumped reaction. This procedure should be repeated until all rows in the stoichiometric matrix have more than two non-zero stoichiometric coefficients, i.e. all reactions of linear pathways are merged. However, attention needs to be paid that metabolic processes, which include metabolite transports, are not lumped, since this lumping would disturb the charge and proton balances. Note, metabolic processes, for which standard Gibbs energies of reactions can be estimated from component-contribution method, can be excluded from being lumped. The exclusion of these metabolic processes ensures that the second law is also considered locally for these metabolic processes, and therefore metabolites being part of these metabolic processes can be additionally constrained.

(iii) Third, linear dependent metabolic processes are identified by flux coupling analysis (96)

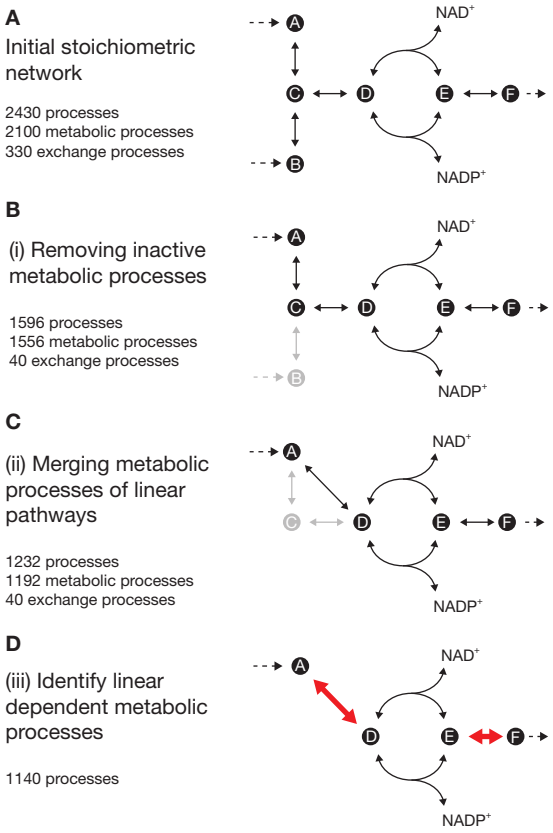

Figure 3. The automated model extraction and reduction is performed in multiple steps. **(A)** Initial stoichiometric network. **(B)** Inactive processes are removed by flux variability analysis. **(C)** Processes of linear pathways are detected and merged. **(D)** Fully coupled processes are identified by flux coupling analysis.

(Fig. 3D), because such coupling allows the reduction of the model complexity (i.e. reduction in the binary variables), and accomplish a reduction in the number of optimizations that need to be performed in the later flux variability analyses. To this end, we determine the minimum and maximum flux ratios between any pair of two metabolic or exchange processes. Two processes are identified as linear dependent fully coupled, when their minimum and maximum flux ratios are identical. These linear dependent fully coupled processes are then grouped into subsets of the metabolic processes, where the each of this subsets is a degree of freedom that is later analyzed.

For building the thermodynamic metabolic network model of *E. coli*, we reduced the model considering the following substrates: acetate, fructose, galactose, glycerol, glucose, mannose, pyruvate, succinate and xylose. Using flux variability analysis 544 metabolic processes (294 chemical conversions, 250 metabolite transport) and 290 exchange processes were identified as inactive under the considered conditions and thus removed from the network. By lumping reactions of linear pathways, we further reduced the network by 339 chemical conversions. Therefore we reduced the model from 2075 to 1192 metabolic processes and from 330 to 40 exchange processes (total reduction of 48 %). By flux coupling analysis, we further reduced the 1232 processes (1192 metabolic- & 40 exchange-processes) to 1140 processes.

Step IV. Training the model

Fourth, unknown parameters of the thermodynamic metabolic model are estimated from experimental training data. These model parameters are (i) a thermodynamic consistent set of standard Gibbs energies for the chemical conversions (incl. chemical conversions for which no estimate can be obtained from the component contribution method), (ii) narrower physiological ranges for the concentrations and for the standard Gibbs energies of reactions, and (iii) the maximal cellular rate of entropy production.

(i) The training is done by regression analysis according to Chapter 2, methods 6, which also includes parametric bootstrap to determine the accuracy of the parameter estimates. The training data consists of measurements, i.e. mean and standard errors, of extracellular rates, the estimates for the standard Gibbs energies from the component contribution method (including mean and standard errors, see step II), and measurements of intracellular metabolite concentrations. The extracellular rates and metabolite concentrations should be measured from exponentially growing cultures at different conditions. The cultures should be selected in such a way that the cells grow at different growth rates (incl. unlimited growth) and have different metabolic operations, i.e. purely respiratory and respiro-fermentative metabolism, and therefore the model parameters, e.g. the maximum entropy production rate, contain the differences in the metabolic operations. To obtain a wide range of different growth rates, chemostat cultures are ideally used.

(ii) After the regression analysis, we validate the estimated parameters by predicting metabolic operations for different growth conditions. These predictions are made by flux balance analysis with a cellular objective of maximizing growth according to Chapter 2, methods 8. The predicted metabolic operations are then evaluated by comparison with test data, i.e. extracellular rates obtained from conditions that were not used for the training.

(iii) The solution space of the validated model is then further statistically analyzed using variability analysis and Markov Chain Monte Carlo (MCMC) sampling according to Chapter 2, methods 8. The variability analysis and MCMC sampling allows to exactly predict different intracellular quantities, e.g. intracellular fluxes, concentrations and entropy production rates.

Discussion

The developed workflow allows us to convert genome-scale metabolic reconstructions, which exist for a large range of different organisms, into a thermodynamic metabolic network model. Such a thermodynamic metabolic network model allows to predict and analyze intracellular fluxes without relying on heuristic assumptions. To build such a model, one requires only a carefully selected set of easily obtainable training data, i.e. extracellular rates, standard Gibbs energies of reactions, and intracellular metabolite concentrations. Key for the development of such models are (i) a correct thermodynamic description and (ii) a manageable size because the computational analyses are computationally very demanding. Here, we illustrated how such a model can be developed.

The limitation of the thermodynamic metabolic network models is the computational complexity introduced by the nonlinearity of these models. Therefore, in order to utilize these models, proficiency with mathematical optimization and high performance computing is re-

quired. This complexity along with the size of the genome scale metabolic network model is the reason, why we only exemplify in this chapter the necessary steps to build such a model.

Methods

Thermodynamic metabolic model

The thermodynamic metabolic model (Chapter 2, methods 5), $M(v,\ln c) \leq 0$, consists of a set of linear and non-linear equations, which are based on: the mass balances including charge and proton balances; the cellular Gibbs energy balance; an equation to calculate the cellular entropy production rate σ^{cell}; equations to calculate the entropy production rates σ; equations to calculate the Gibbs energy exchange rates g; equations to calculate the Gibbs energies of reactions $\Delta_r G'$; equations to calculate the Gibbs energies of formation $\Delta_f G'$ of the metabolites i that are transferred across the system boundary; the second law of thermodynamics for the metabolic processes and the growth process,

$$\left\{M(v,\ln c) \leq 0\right\} \triangleq \left\{ \begin{array}{c} \sum_{j \in MET} S_{ij} v_j = v_{i \in EXG} \quad \forall i \\ -T\sigma^{cell} = \sum_{i \in EXG} g_i \\ \sigma^{cell} = \sum_{j \in MET} \sigma_j \\ \sigma_j = -\frac{\Delta_r G'_j v_j}{T} \quad \forall j \in MET \\ g_i = \Delta_f G'_i v_i \quad \forall i \in EXG \\ \Delta_r G'_j = \Delta_r G'^{o}_j + \Delta_r G'^{t}_j + RT \sum_{i \notin h^+} S_{ij} \ln c_i \quad \forall j \in MET \\ \Delta_f G'_i = \Delta_f G'^{o}_i + RT \ln c_i \quad i \in EXG \\ \sigma_j > 0 \quad \left|\Delta_r G'_j\right| \geq 0.5 \quad \forall j \in MET \setminus \{BMSYN, H2Ot\} \\ \sigma_{BMSYN} + \sigma_{ATPH} \geq 0 \end{array} \right\} \quad , \text{ (Eq.1)}$$

where: i are the metabolites, $j \in MET$ are the metabolic processes, $i \in EXG$ are the exchange processes, S is the stoichiometric matrix, v are rates through metabolic and exchange processes, T is temperature, $\Delta_r G'^{o}$ are standard Gibbs energies of the chemical conversions, $\Delta_r G'^{t}$ are the Gibbs energies of metabolite transports, R is the universal gas constant, ln c are the metabolite concentrations on the logarithmic scale, *BMSYN* is the biomass synthesis reaction, *H2Ot* are the water transporters across the membranes, *ATPH* is the ATP hydrolysis reaction.

Determining chemical species and Gibbs free energies of formation using the component contribution method

A list of formulas of all metabolic processes, in which the metabolites were encoded by their KEGG ID's, was used as input for the component contribution method. Metabolites, for which no unique KEGG ID exists, were given a consecutive 'fake KEGG ID' starting from C99999 backwards.

In a first step, the CCM determines for each metabolite, calling ChemAxon, the pKa values

and subsequently all at pH 7 present species with charge and number of protons. For each species, the Gibbs free energy of formation is then estimated using a combination of group and reactant contribution method. Finally, the standard deviation for each Gibbs free energy of reaction is calculated using a cross-validation benchmark and a set of measured Gibbs free energies of reactions.

Flux variability analysis and flux coupling analysis

For the reduction of the stoichiometric network (Step III), we formulated a solution space Ω^{MR} based on the linear constraints of $M(v,\ln c) \leq 0$ (Eq. 1),

$$\Omega^{MR} = \left\{ v \mid \left(\sum_{j \in MET} S_{ij} v_j = v_{i \in EXG} \quad \forall i \right) \wedge \left(v_i^{lo} \leq v_i \leq v_i^{up} \ \forall i \in EXG \right) \right\} \quad , \quad \text{(Eq.2)}$$

where we constrained the extracellular rates, $i \in EXG$ according to the medium conditions. Note: Because the solution space Ω^{MR} only consists of linear constrains, we use linear programming algorithms (CPLEX) to solve the optimizations. To determine the flux variability (16), we solved the following optimization problems,

$$v_{j(i)}^{\min/\max} = \min/\max \left\{ v_{j(i)} : v \in \Omega^{MR} \right\} \quad \forall j \in MET \vee i \in EXG \quad , \quad \text{(Eq.3)}$$

where we minimized and maximized v of every metabolic and exchange process. A metabolic or exchange process, for which the minimum and maximum rate was zero, was identified as inactive and subsequently removed from the model. To determine linear dependent reactions, we determined the minimum and maximum flux ratios,

$$R_{j(i)1\,j(i)2}^{\min/\max} = \min/\max \left\{ v_{j(i)1} / v_{j(i)2} : v \in \Omega^{MR} \right\} \quad \forall j1, j2 \in MET \vee i1, i2 \in EXG \quad , \quad \text{(Eq.4)}$$

for each pair *j1,j2 (i1,i2)* of two metabolic or exchange processes. We then identified two processes as linearly dependent when the minimum equaled the maximum flux ratio according to (96).